View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by CERN Document Server

arXiv:physics/0703208v1 22 Mar 2007

Statistical properties of short term price trends in high frequency stock market data

Paweł Sieczka, Janusz A. Hołyst *

Faculty of Physics, Center of Excellence for Complex Systems Research, Warsaw University of Technology, Koszykowa 75, PL-00-662 Warsaw, Poland

Abstract

We investigated distributions of short term price trends for high frequency stock market data. A number of trends as a function of their lengths was measured. We found that such a distribution does not fit to results following from an uncorrelated stochastic process. We proposed a simple model with a memory that gives a qualitative agreement with real data.

Key words: Econophysics, Financial markets, Price trends
PACS: 89.65.Gh

1 Introduction

Statistical analysis of stock prices is a rich source of information about the nature of financial markets. It was Louis Bachelier who as the first used a stochastic approach to model financial time series [1]. Since that time the analysis has become a widely investigated area of interdisciplinary researches [2,3,4].

In 1973, Fischer Black and Myron Scholes published their famous work [5] where they presented a model for pricing European options. They assumed that a price of an asset can be described by a geometric Brownian motion. However, the behavior of real markets differs from the Brownian property [6,7], since the price returns form a truncated Lévy distribution [8,9,10]. As a result of this observation many non-Gaussian models were introduced [11,12,13].

* Corresponding author. Tel.: +48 22 234 71 33; fax: +48 22 234 58 08
Email address: jholyst@if.pw.edu.pl (Janusz A. Hołyst).

Another divergence from Gaussian behavior is an autocorrelation in financial systems. Empirical studies show that the autocorrelation function of the stock market time series decays exponentially with a characteristic time of a few minutes, while the autocorrelation of prices absolute values decays slower, as a power law function, what leads to a volatility clustering[14,15,16,17].

It has been observed [18] that for certain time scales a sequence of two positive price changes leads more frequently to a subsequent positive change than a sequence of mixed changes, i.e. the conditional probability $P(+\backslash++)$ is larger than $P(+\backslash+-)$. We will investigate this effect hereafter.

2 Empirical data

In this paper we study short term price trends for high frequency stock market data. By *short term uptrend/downtrend* we mean such a sequence of prices that a price is larger/smaller than the preceding one (see below for a more precise definition).

First, having a time series Y_t, which is in our case a history of a stock price or a market index, we build a series of variables s_t in the following way:

- $s_t = 1$ if $Y_t > Y_{t-1}$,
- $s_t = -1$ if $Y_t < Y_{t-1}$,
- $s_t = s_{t-1}$ if $Y_t = Y_{t-1}$.

A positive value of the variable s_t means that at the time t the price Y_t did not decrease, and similarly a negative value means that the price did not increase.

In a series s_t we can distinguish subseries of identical values. For $a < b$ and $s_a = s_b = s$, $S(a, b, s)$ is such a subseries if and only if $\forall_{c\in(a,b)}\, s_c = s$. Subseries $S(a, b, s)$ can be identified with an uptrend lasting from $t = a$ till $t = b$ for $s = 1$ and with a downtrend for $s = -1$. The length l of such a uptrend/downtrend equals to $b - a + 1$. Let us mention that a subseries of a length l includes two subseries of length $l - 1$, three subseries of length $l - 2$ etc.

Let $N(l)$ be a number of subseries of a length l with a fixed s in a series $s_1, ..., s_M$. If s_t were generated by an uncorrelated discrete stochastic process with a probability $P(s_t = 1) = p$, then the expected value of $N(l)$ would be equal to:

$$N(l) = (M - l + 1)p^l, \tag{1}$$

where M is a number of all elements in the basic series. Similarly the expected value of downtrend series of length l is $N(l) = (M - l + 1)(1 - p)^l$.

We have measured the distribution $N(l)$ for real market data and the same

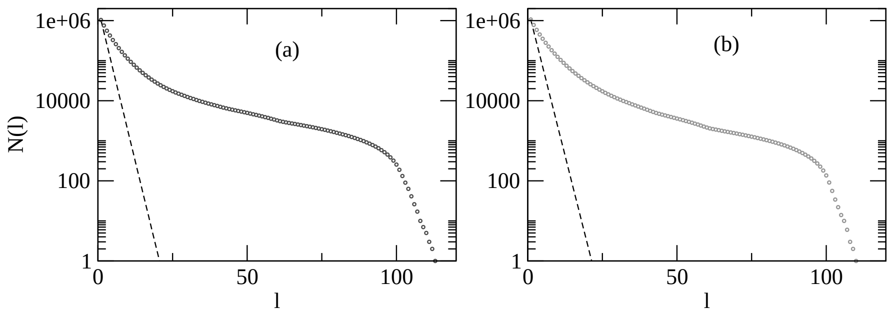

Fig. 1. Distribution $N(l)$ for uptrends (a) and downtrends (b) for the WIG20 index sampled every 15 seconds (circles), compared to uncorrelated process (line) eq. (1).

distribution for the corresponding uncorrelated process. Figure 1 presents this distribution for the WIG20 index of Warsaw Stock Exchange (WSE) and the same distribution for the corresponding uncorrelated process. The results show a significant difference between real data and the uncorrelated model. If variables s_t were uncorrelated, there would not be subseries longer than 25 ticks. In fact, subseries even longer than 100 ticks are present. The trends last for about 30 minutes. There are far more such trends than it would be if the process were uncorrelated. The distribution $N(l)$ was also calculated for particular stocks from WSE, NYSE and NASDAQ (fig. 2).

The observed difference between the uncorrelated model and the real markets is due to strong autocorrelations in the process s_t. It is only seen in high frequency data. Choosing every n-th element of the series s_t weakens the autocorrelations and makes the outcome approaching the uncorrelated model with growing n. It is shown in fig. 3.

3 A phenomenological model of correlated market prices

In real markets, variables s_t and $s_{t+\tau}$ are correlated although this correlations decay very fast. Let $r(k)$ stand for a conditional probability $P(s_{n+k+1} = 1 \backslash s_{n+k} = 1, ..., s_n = 1, s_{n-1} = -1)$, which is independent of n. For processes where autocorrelations are present we can write a generalization of equation (1):

$$N(l+1) = (M-l)p \prod_{i=1}^{l} r(i), \tag{2}$$

for $l > 0$ and $N(1) = Mp$.

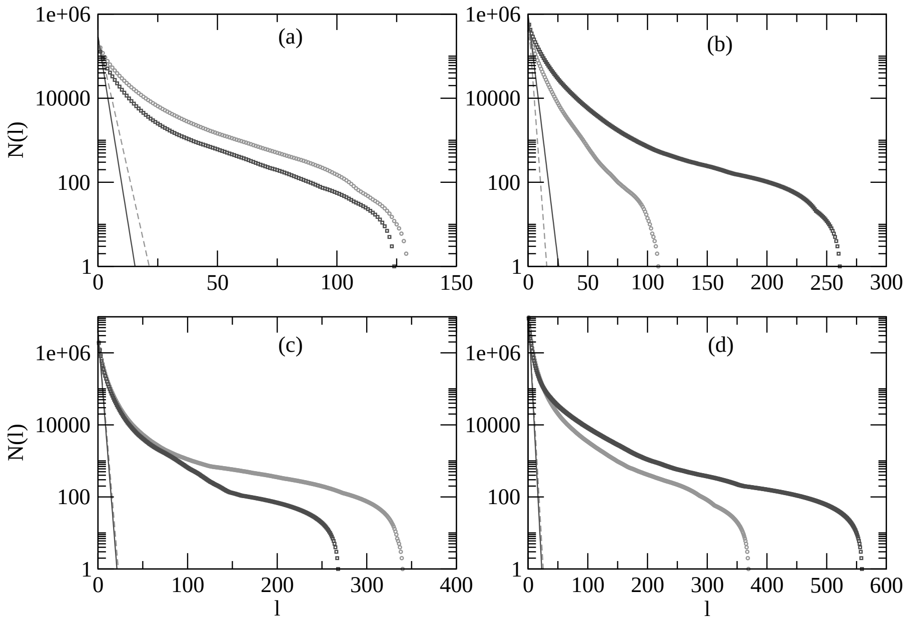

Fig. 2. Distribution $N(l)$ for: (a) BIOTON (WSE), (b) TPSA (WSE), (c) APPLE (NYSE), (d) INTEL (NASDAQ). Uptrends are plotted with circles and downtrends are plotted with squares. All data are sampled tick by tick. Lines correspond to the uncorrelated process (1).

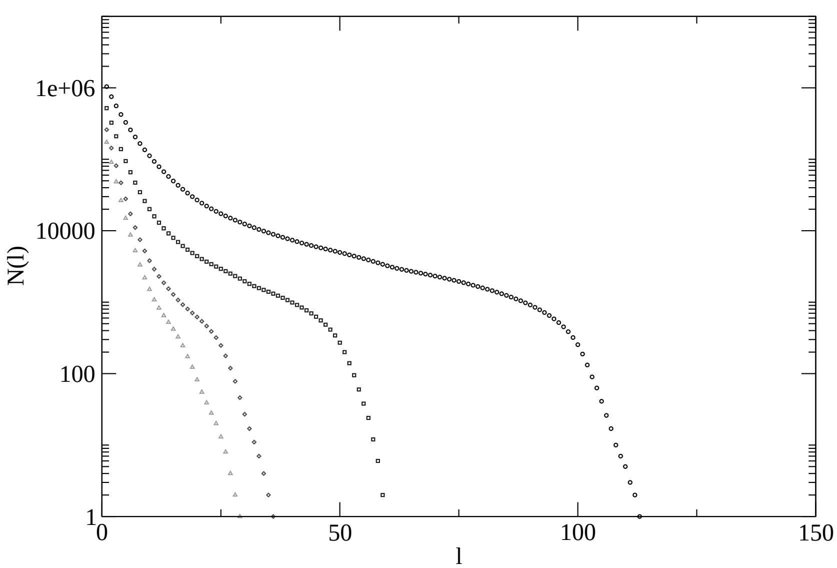

Fig. 3. $N(l)$ of WIG20 uptrends for every n-th element of s_t, n=1 circles, n=2 squares, n=4 diamonds, n=6 triangles.

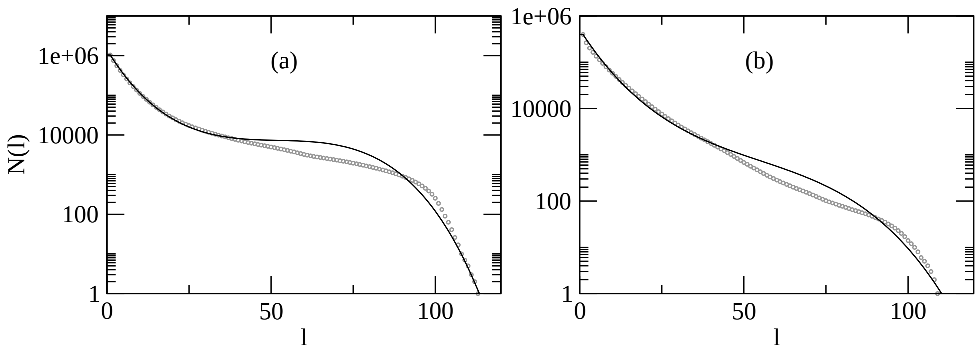

Fig. 4. The distribution $N(l)$ with fitted curve (2). (a) WIG20 (WSE) uptrends with fitted parameters: $a = -0.000098$, $x_1 = -48.28$, $x_2 = 153.31$, (b) TP S.A. company (WSE) uptrends with fitted parameters: $a = -0.000057$, $x_1 = -73.42$, $x_2 = 184.35$.

Let us see that the result (2) is equivalent to (1) if for any x there is $r(x) = p$. A key issue is to model $r(x)$ in order to describe characteristics of a given market. The results presented in fig. 1-2 show that the model of the uncorrelated process (1), is a poor simplification. To get a better consistency with real data the function $r(x)$ can be modelled as:

$$r(x) = a(x - x_1)(x - x_2), \tag{3}$$

with fitted parameters a, x_1 and x_2. A binomial function was chosen because we are looking for a simple concave function with a maximum, and a binomial function matches our requirements for proper parameters a, x_1, x_2.

We expect that for small x, the probability value $r(x)$ increases with x. It means that when the trend starts forming, investors follow it and, as a result, they amplify the trend. Thus, the probability of a continuation of the price movement grows. As time goes by, some of them may want to withdraw to take profits and those who are out of the market believe it is to late to get in. This causes a decrease of $r(x)$ for a longer trend.

One can choose various functions to model the probability $r(x)$. All such functions should be concave functions with a maximum for a positive argument smaller than the maximal length of all subseries. The figure 4 presents the distribution $N(l)$ with a fitted curve based on (3).

Putting (3) into (2) we get after some algebra an approximated form of the function $N(l)$ as

$$\begin{aligned} N(l) \simeq{} & (M - l + 1)p\mathrm{e}^{-2(l-1)}[a(l - x_1)(l - x_2)]^{l-1} \\ & \times\left(\frac{l - x_1}{1 - x_1}\right)^{1-x_1}\left(\frac{l - x_2}{1 - x_2}\right)^{1-x_2}. \end{aligned} \tag{4}$$

The figure 5 presents functions (2) and (4) and a relative difference between them.

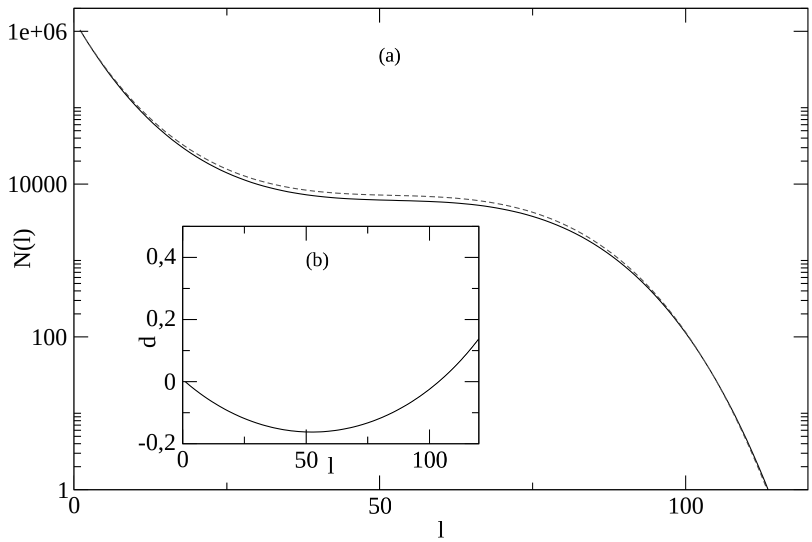

Fig. 5. Graph (a) presents the distribution $N_1(l)$ obtained from the equation (2) and $N_2(l)$ from the approximation (4) for the WIG20 index (WSE). Graph (b) presents the relative difference between them: $d = (N_2(l) - N_1(l))/N_2(l)$.

For a given set of parameters $a^+ = -0.000098$, $x_1^+ = -48.28$, $x_2^+ = 153.31$, $a^- = -0.000101$, $x_1^- = -47.32$, $x_2^- = 150.69$, obtained for uptrends and downtrends in the WIG20 index respectively, one can simulate the stochastic process according to eq. (2, 3). The autocorrelation function

$$C(\tau) = \langle s_{i+\tau} s_i \rangle, \tag{5}$$

calculated for such a process decrease similarly to the autocorrelation function received from empirical data see fig. 6. The model reflects short range correlations of the sign.

4 Conclusions

We have investigated short term price trends for high frequency stock market data. It turned out that the statistics for real markets is significantly different from the statistics of uncorrelated processes. Longer trends (of the order of several minutes) are much more frequent than they should be, if one used an uncorrelated model.

We proposed a simple model that qualitatively captures the behavior of the market. Our model leads to a distribution of trend series $N(l)$ that is similar to the distribution observed in market data. Our model produces also short range correlations. This behavior is caused by the conditional probability of

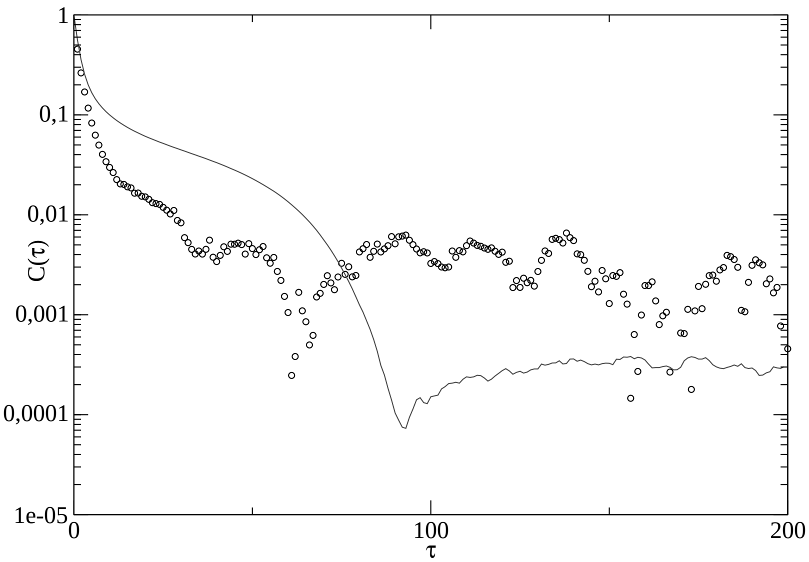

Fig. 6. The autocorrelation function obtained from eq. (5) for parameters $a^{+} = -0.000098$, $x_1^{+} = -48.28$, $x_2^{+} = 153.31$, $a^{-} = -0.000101$, $x_1^{-} = -47.32$, $x_2^{-} = 150.69$ (line) and the index WIG20 itself (circles).

trend continuation that changes nonmonotonically with a trend length. At the beginning of the trend, the probability of the trend continuation grows, then it hits the maximum and finally decreases. As a result trends posses a limited length.

Acknowledgments

This work as a project of the COST Action P10 "Physics of Risk" was partially supported by Polish Ministry of Science and Higher Education, Grant No. 134/E-365/SPB/COST/KN/DWM 105/2005-2007.

References

[1] L. Bachelier, Theorie de la speculation, Ann. Sci. Ec. Norm. Super. 111117 (1890) 2186.

[2] K. Urbanowicz, J. A. Hołyst, Phys. Rev. E 67 (2003) 046218

[3] Proceedings from the 4-th International Conference APFA4, Phys. A 344 (2004) Issue 1-2.

[4] Trends in Econophysics, Eur. Phys. J. B 55 (2007) 121-217.

[5] F. Black, M. Scholes, J. Political Economy 81 (1973) 637-654 .

[6] B. Mandelbrot, J. Business 36 (1963) 394419.

[7] E. F. Fama,, J. Business 38 (1965) 34105.

[8] R. N. Mantegna, H. E. Stanley, Phys. Rev. Lett. 73 (1994) 2946.

[9] R. N. Mantegna, H. E. Stanley, Phys. A 254 (1998) 77-84.

[10] L. Kullmann, J. Töyli, J. Kertesz, A. Kanto, K. Kaski, Phys. A 269 (1999) 98-110.

[11] L. Borland, Phys. Rev. Lett. 89 (2002) 098701.

[12] S. G. Kou, Manage. Sci. 48 (2002) 1086-1101.

[13] S. G. Kou, Manage. Sci. 50 (2004) 1178-1192.

[14] Y. Liu, P. Cizeau, M. Meyer, C.-K. Peng, H. E. Stanley, Phys. A 245 (1997) 437-440.

[15] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, H. E. Stanley, Phys. Rev. E 60, 1390.

[16] P. Gopikrishnan, V. Plerou, L. A. N. Amaral, M. Meyer, H. E. Stanley, Phys. Rev. E 60, 5305.

[17] A. Krawiecki, J. A. Hołyst, D. Helbing, Phys. Rev. Lett. 89, 158701.

[18] E. F. Fama, J. Finance 25 (1970) 383.